

4.2. Information Model

This section builds on the concepts presented in section 2 to further describe the types of information that are exchanged and managed within the OAIS. This section also defines the specific information objects which are used within the OAIS to preserve and access the information entrusted to the archive. This more detailed model of OAIS related information structures is intended to aid the architect or designer of future OAIS systems. The structures discussed in this section are conceptual and should not be taken to imply any specific implementations.

As discussed in Section 2, the primary goal of an OAIS is to preserve information for a designated community over a indefinite period of time. In order to preserve this information an OAIS must store significantly more than the contents of the object it is expected to preserve. This section analyzes those information requirements to describe the object classes of data associated with an OAIS. This section uses “**Object Modeling Technique** (OMT) -like” [Reference 2] diagrams to model the concepts discussed in the text. An overview of the notation used and critical object modeling concepts is presented in Annex D of this document. Since the all the objects in this model are information structures, the dynamic modeling aspects of OMT have not been used.

Section 4.2.1 provides a model of the information required for effective long term preservation of information while Section 4.2.2 describes the conceptual objects and containers that represent the contents of an OAIS.

4.2.1 Logical Model for Archival Information

4.2.1.1 Information Object

A basic concept of the OAIS Reference Model is the concept of information being a combination of data and Representation Information. This concept is illustrated by the OMT diagram 4-5. The **Information Object** (IO) is composed of a Data Object which is either physical or digital and the Representation Information that allows for the full interpretation of the data into meaningful information. This model is valid for all the types of information in an OAIS.

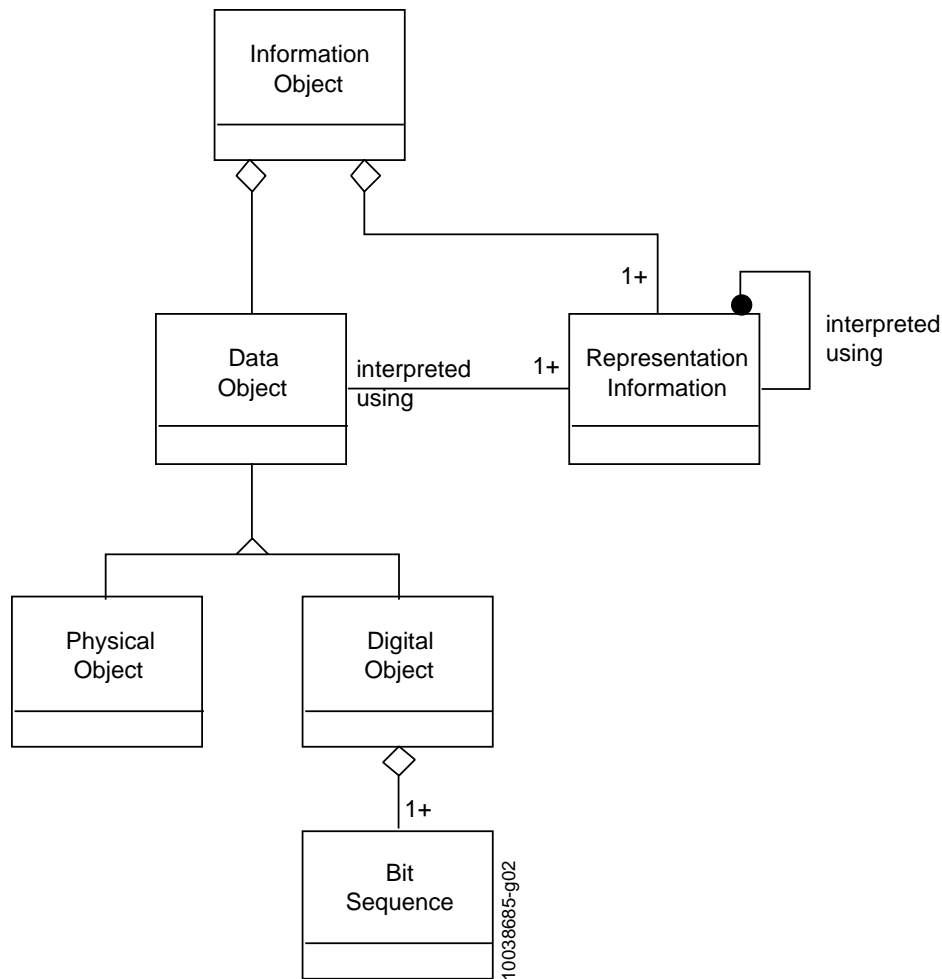


Figure 4-5. Information Object

4.2.1.2 Taxonomy of Information Object Classes

There are many types of information that are needed for the long-term preservation of information in an OAIS. Each of these types can be viewed as a complete Information Object in that it contains a data object and adequate representation information to understand the data. This section builds on the discussions in section 2.2 about the types of supporting information needed to enable long term preservation and the role of Representation Information. The information modeling in this section discusses several types of types of Information Objects that are used in the OAIS. The objects are categorized by their content and function in the operation of an OAIS into content objects, preservation description objects, packaging information objects, and descriptor objects. The following sections discuss the contents of each of the types of Information Object. Figure 4-6 shows a taxonomy of these information objects..

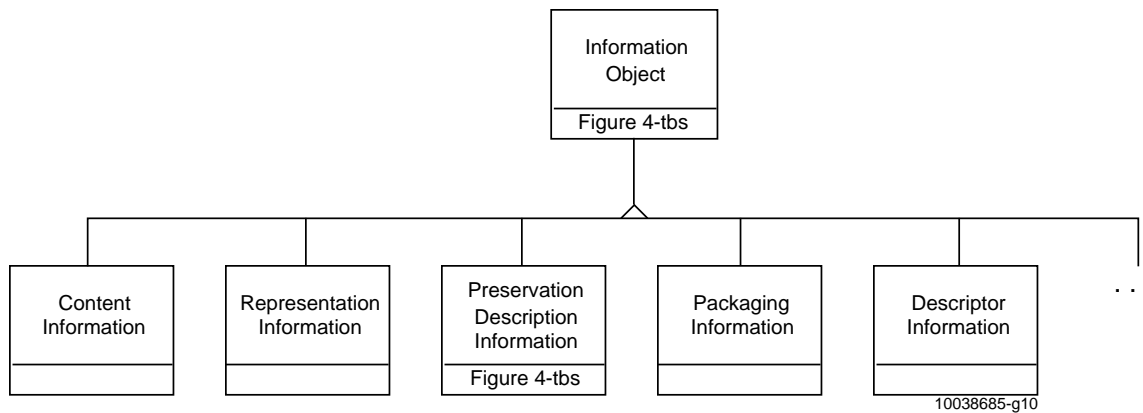


Figure 4-6. Information Object Taxonomy

4.2.1.2.1 Content Information

The **Content Information** is that information which is the primary object of preservation.. The Content Information can be viewed as a primary **Data Object** together with its Representation Information as shown in Figure 4-5 . The Data Object in the Content Information may be either a Digital Object or a Physical Object (e.g., a physical sample, microfilm). Any Information Object may serve as Content Information. The special thing about an instance of Content Information is that it is the information that an archive is tasked to preserve.

4.2.1.2.2 Representation Information

The CI may be expressed as either a physical object (e.g., a moon rock) together with some **Representation Information**, or it may be expressed as a digital object (i.e., a sequence of bits) together with the Representation Information giving meaning to those bits.

The Representation Information accompanying a physical object like a moon rock may give additional meaning, as a result of some analysis, to the physically observable attributes of the rock. This information may have been developed over time and the results, if provided, would be part of the CI.

The Representation Information accompanying a digital object, or sequence of bits, is used to provide additional meaning. It typically maps the bits into commonly recognized data types such as character, integer, and real and into groups of these data types. It may associate these with higher level meanings which can have complex inter-relationships that are also described.

The remainder of this section focuses on the Representation Information object when the Data Object is specialized as a Digital Object.

The Digital Object, as shown in Figure 4-5, is itself composed of one or more bit sequences. The purpose of the Representation Information object is to convert the bit sequences into more meaningful information. It does this by describing the format, or data structures, which are to be applied to the bit sequences and that in turn result in more meaningful values such as characters, numbers, pixels, arrays, tables, etc. These common computer data types, and aggregations of these data types, are referred to as the Structure

Layer information of the Representation Information object. These structures are commonly identified by name or by relative position within the associated bit sequences.

The Representation Information provided by the Structure Layer is seldom sufficient. Even in the case where the Digital Object is interpreted as a sequence of text characters, and described as such in the Structure Layer, the additional information as to which language was being expressed should be provided. This higher layer information is referred to as the Semantic Layer, although in reality each layer provides its own set of semantics. When dealing with scientific data, for example, the information in the Semantic Layer can be quite varied and complex. It will include special meanings associated with all the elements of the Structural Layer, and their inter-relationships. An expansion of the Representation Information object is given in Figure 4-7.

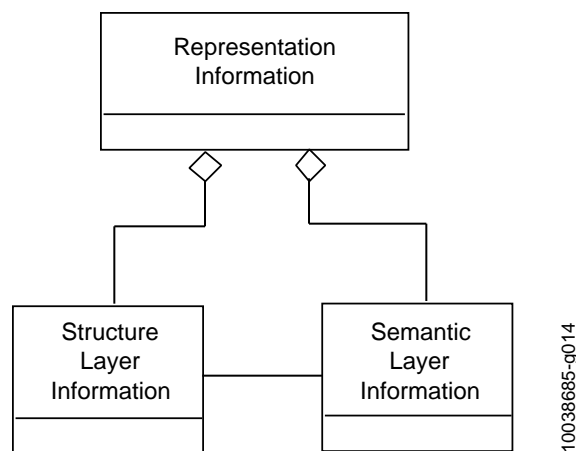


Figure 4-7. Representation Information Object

The Semantic Layer can also be viewed as the "Object" layer. Taking an object oriented view of the combination of the Representation Information and its Digital Object (i.e., Content Information object), queries applied to this object are addressed to the Semantic Layer. The object oriented methods translate these queries into actions on the Structure Layer elements, and ultimately to the bit sequences, with results reported back in Semantic Layer terms. Such software methods associated with a Content Information object provide useful services as long as the software executes properly. However for indefinite long-term information preservation, a full and understandable description of the Representation Information is essential. This can be a particular challenge for the preservation of scientific type data as there are few standards for how to express this type of information and archives need to ensure this information is understandable to the designated consumer communities.

All this additional Representation Information (both semantic and syntactic) is needed to fully transform the bits of a file into the Content Information. In principal, this even extends to the inclusion of definitions (e.g., dictionary and grammar) of the natural language used (English in this example). Over long time periods the meaning of natural language expressions can evolve significantly in both general and in specific discipline usage. As an aside, it is clear from this example that, in general, 100% information preservation for the indefinite long-term is a goal that is not practically achievable.

Representation Information may be expressed in physical forms (e.g., a paper document) or in digital forms. When the Representation Information is in digital form, additional Representation Information is needed to understand the bits of the Representation Information. In principle, this recursion continues until physical forms are encountered. For example, Representation Information expressed in ASCII needs the additional Representation Information for ASCII, which may be a physical document giving the ASCII standard. Because each Representation can be composed of multiple components, each with its own Representation, the result can be described as a **Representation Net**. In practice, the recursive chain of the representation net is also broken when there is widely available software that understands a particular representation, such as ASCII display software. Nevertheless, this situation is dynamic as the technology and standards evolve and is a preservation issue needing continuing attention.

As a practical matter, the OAIS needs to have enough Representation Information associated with the bits of the Data Object in the Content Information that it feels confident that it, and those expected to use the Content Information, can enter the Representation Net with enough knowledge to begin accurately interpreting the Representation Information. This is a significant risk area for an OAIS, particularly for those with a narrow discipline focus, because jargon and apparently widely understood terms may subsequently be found to be quite temporary.

As a complex example, consider an electronic file containing a sequence of values obtained from a sensor looking at the Earth's environment. There is a second file, encoded using ASCII, that provides information on how to understand the first file. It describes how to interpret the bits of the first file to obtain meaningful numbers, it describes what these numbers mean in terms of the physics of the observation being conducted, it gives the date and time period over which the observations were made, it gives an average value for the observed values, and it describes who made the observations. These two files are submitted to an OAIS for preservation.

Assume that the OAIS determines that the Content Information to be preserved is the observed bits together with their values as numbers and the physics meaning of these numbers. This information is conveyed by the bit sequence within the first file together with the **Representation Information** from the second file needed to transform the first file's bits into meaningful physical values. Note that neither the first file's underlying media nor the particular file system carrying the bits is part of the Content Information. From the second file only part of its content is considered a part of the Content Information and this is the part that enables the transformation of the bits from the first file into meaningful physical values. In fact this second file does not carry all the Representation Information needed to make this transformation because the following additional information is needed:

- Information that the second file is encoded in ASCII so that it can be read as meaningful characters;
- Information on how the characters are used to express the transformations from bits to numbers to meaningful physics values. This information, typically referred to as a combination of format information and data dictionary information, may also include instrument calibration values and information on how the calibrations are to be applied. All this information may be widely understandable once the ASCII characters are visible because it has all been expressed in English (or some other natural language), or some of it may be in more structured forms that will need additional Representation Information to be understood.

Therefore the Representation Information of the second file needs additional Representation Information, and this information may need additional Representation Information, etc., forming a linked set of Representations of Representations. This is a good example of the complex Representation Net.

Recall that in the example above, there was a determination that the Content Information consisted of the observed sensor values and their meanings. This is by no means the only determination that could have been made. It could just as easily have been determined that the Digital Object of the desired Content Information was the bit sequences within the first file together with the all the bit sequences within the second file. The fact that some of these latter bit sequences are used to interpret the first files bit sequences is just an example of a set of bits that is somewhat self-describing. It is irrelevant that some of the bits in the second file are the basis for information on the date and time period over which the observations were made, the average value for the observed values, and who made the observations. Once it has been determined that all these bits constitute the Digital Object of the Content Information, then the Representation Information is that information needed to turn them into meaningful information. How extensive this meaning is to be carried and how far the Representation Net needs to be carried are local issues for the OAIS and its related producer and consumer communities.

The extent to which the CI is understandable to a designated community depends largely on the nature of the Representation Information. It needs to be written using constructs which are understandable to that community and it needs to be sufficiently complete to convey all the information intended. Narrowly drawn or specialized communities may need minimal Reference Information to understand a particular CI, but in such cases extra care needs to be exercised to ensure that the natural evolution of what is commonly understood in those communities does not effectively cause information loss from the CI.

4.2.1.2.3 Preservation Description Information

In addition to the content object the Information Package must include a set of IOs which will allow the understanding of the content objects over an indefinite period of time. The specific set of IOs which are required for this function are called collectively called **Preservation Description Information (PDI)**. The PDI IOs are a specialization of the IO shown in figure 4-5 in that the data object in a PDI class must be digital while the data object in an IO may be a digital object or a physical object. The purpose of this restriction is to emphasize the fact that in modern archives even those with physical CI the PDI will tend to be digital and need the representation information defined that is described in the previous subsection. PDI is that information which is necessary to adequately preserve the particular Content Information with which it is associated. It is specifically focused on describing the past and present states of the Content Information, ensuring it is uniquely identifiable, and ensuring it has not been unknowingly altered. This information is typical for all types of archives and has been classified in the context of traditional archives. However, the class definitions must be extended for digital archives. The following definitions are based on the categories discussed in the paper “Preserving Digital Information”. Figure 4-8 in an OMT diagram which illustrates the taxonomy of PDI types.

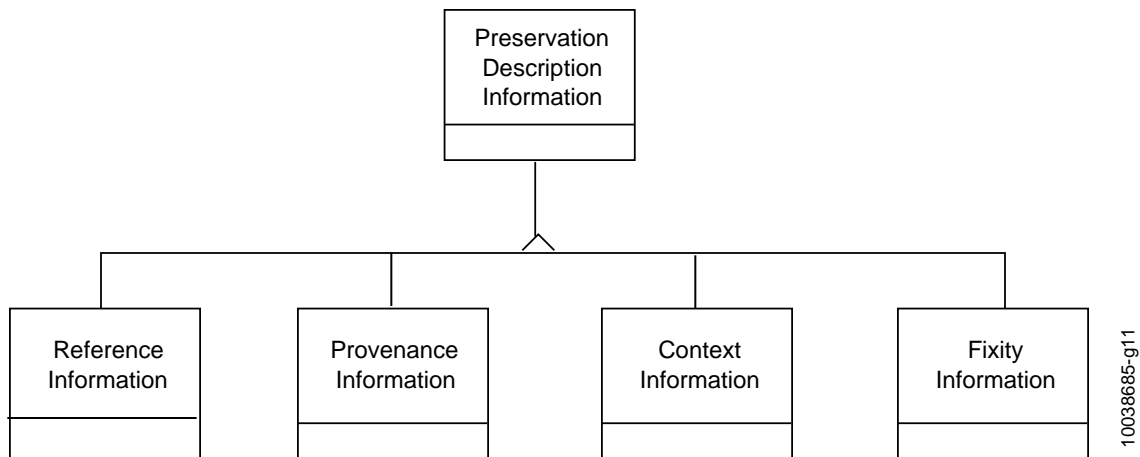


Figure 4-8. Preservation Description Information

- **Provenance Information:** This information documents the history of the Content Information. This tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. This give future users some assurance as to the likely reliability of the Content Information.
- **Reference Information:** This information identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the Content Information. It also provides those identifiers that allow outside systems to refer, unambiguously, to this particular Content Information.
- **Context Information:** This information documents the relationships of the Content Information to its environment. This includes why the Content Information was created, and how it relates to other Content Information objects existing elsewhere.
- **Fixity Information:** This information documents the authentication mechanisms, and it provides any authentication keys used to ensure that the particular Content Information object has not been altered in an undocumented manner.

The OAIS needs to explicitly decide what the Content Information is in order to be able to ensure that it also has the necessary PDI which is needed to preserve the Content Information. Deciding what is the Content Information, for a given set of information, may not be obvious and may need to be negotiated with the information producer. Once the Content Information has been determined, it is possible to assess the Preservation Description Information.

4.2.1.2.4 Packaging Information

The **Packaging Information** is that information which, either actually or logically, binds and relates the components of the package into a physical entity on specific media. This packaging information consists of all the information necessary to delimit the IP and to locate the IP on the media. For example, if the CI and PDI are identified as being the content of specific files on a CD-ROM, then the PI may be the ISO-9660 volume/file structure on the CD-ROM. The PI, in this case, may also include the physical CD-ROM disk. These choices are the subject of local archive definitions or conventions. The

Packaging Information does not necessarily need to be preserved by an OAIS since it does not contribute to the Content Information or the PDI, however, there are many cases where the archive is required to exactly reproduce the submission to the archive. In these cases the Packaging Information must be preserved and reproduced. The OAIS should also avoid hiding PDI or Content Information in the naming conventions in directory or file structures. Since Packaging Information is not preserved this information may be lost unless it is preserved in the CI or PDI. The subject of Packaging Information is an important consideration to the Migration of Information in an OAIS to newer media. This subject is addressed in detail in section 5 of this document.

4.2.1.2.5 Descriptor Information

The Information Objects described previously in this section, provide the information necessary to enable the long term preservation function of the archive. In addition to preserving information, the OAIS must provide adequate features to allow consumers to locate information of potential interest, analyze that information, and order desired information. This is accomplished through a specialization of the Information Object called a **Package Descriptor** which contain the data that serves as the input to documents or applications called **Access Aids**. Access Aids can be used to locate, analyze or order information from the OAIS. The information needed for one access aid is called an **Associated Description**. A single Package Descriptor may contain several associated descriptions depending on the number of different access aids that can locate, visualize, or order the associated AIP.

In addition to the Associated Descriptions, the Descriptor may also contains **Access Methods** which enable authorized users to retrieve the AIU described by the Unit Descriptor from Archival Storage. In most current archives, only internal archive processes and operations personnel and functions are authorized to use these Access Methods. However, as technology advances increases the processing power of the archive and the bandwidth between the archive and the user such access methods as “content based queries” and “data mining” may provide the user with direct read only access to the content objects of AIPs.

The Package Descriptor is not required for the long-term preservation of the content information but is needed to provide visibility and access into the contents of an archive. The contents of the Package Descriptor are highly dependent on the type of AIP it describes. The uses and types of product descriptors in an OAIS are further defined in the section 4.2.2.

4.2.2 Logical Model of Information in an Open Archival Information System (OAIS)

The previous section defines the types of Information Objects that are needed by an OAIS to enable the long term preservation of information and effective access to the preserved information by the designated community. This section uses those IO descriptions to model the conceptual information structures required to accomplish these functions. The models presented in this section are not intended to imply a concrete implementation but rather to highlight the relationship among the types of information needed in the archival process.

4.2.2.1 Information Package (IP)

The conceptual structure for supporting long term preservation of information is the Information Package (IP) . The OMT diagram in figure 4-9 illustrate the conceptual view of an IP. An information package is a container which contains two types of information objects, the Content Information (CI), and the Preservation Description Information (PDI); The IP can be associated with two other types of information objects, Packaging Information (PI) and Package Descriptors.

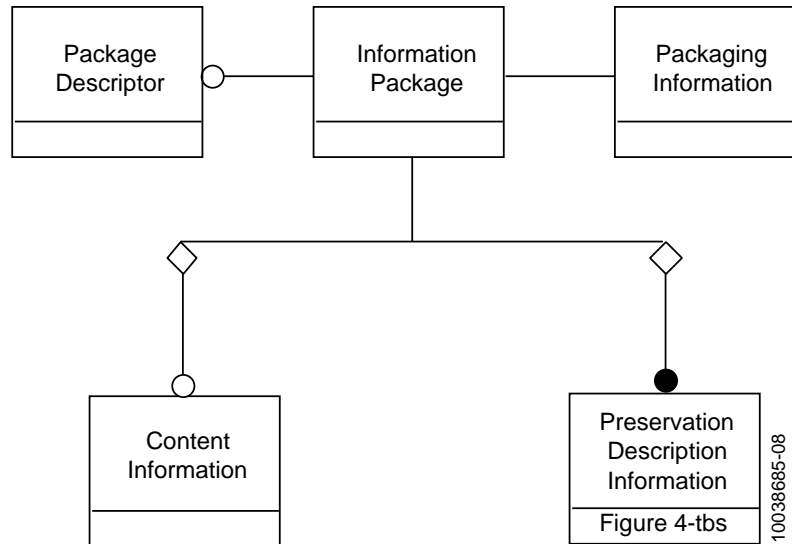


Figure 4-9. Information Package Contents

4.2.2.2 Type of Information Packages

There are three subtypes of the IP identified in section 2.2, the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP). The definitions of these package types in section 2 is based on the function of the archival process which uses the package and the translation from one package to another as it passes through the archival process. This taxonomy of IP types is shown in Figure 4-10.

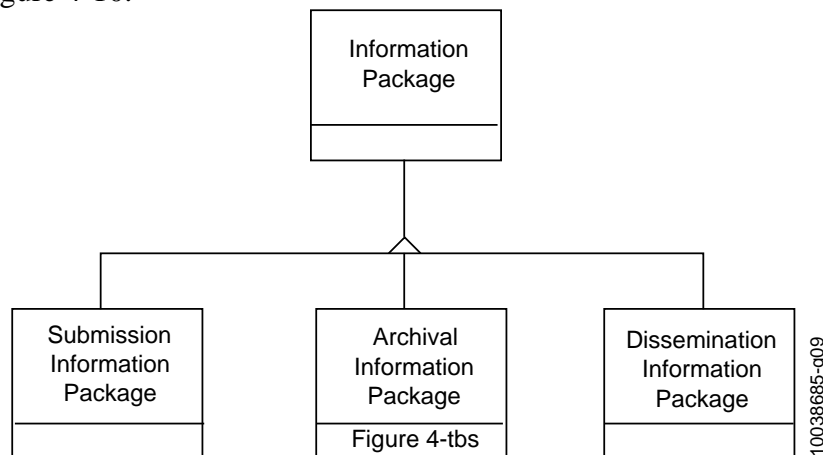


Figure 4-10. Information Package Taxonomy

The **Submission Information Package (SIP)** is that package that is sent to an OAIS by a Producer. Its form and detailed content is typically negotiated between the Producer and the OAIS. Most SIPs will have some CI and some PDI, but it may require several SIPs to provide a complete set of CI and associated PDI. The PI will always be present in some form.

Within the OAIS one or more SIPs is transformed into one or more **Archival Information Packages (AIP)** for preservation. The AIP has a complete set of PDI for the associated CI. The AIP may also contain a collection of other AIPs and this is discussed and modeled in Section 4. The Packaging Information of the AIP will conform to OAIS internal standards, and it may vary as it is managed by the OAIS.

In response to a request, the OAIS provides all or a part of an AIP to a Consumer in the form of a **Dissemination Information Package (DIP)**. The DIP may also include collections of AIPs, and it may or may not have complete PDI. The Packaging Information will always be present in some form so that the Consumer can clearly distinguish the information requested. The Packaging Information may take several forms depending on the dissemination media and Consumer requirements.

Though the implementation of the AIP may vary from archive to archive, the specification of the AIP contains all the needed information to allow long-term preservation and access to archive holdings. The information model for the AIP presented in the section 4.2.2.3. should be used as a reference to establish the types of information required to enable long-term preservation and access.

The exact information contents of the SIP and DIP and their relationship to the corresponding AIP are dependent on the agreements between the archive and its producers and consumers. The model for both of these packages is the same as for the IP with the same cardinality and optionality among contained IOs. The subject of transformations between SIP and AIP, and between AIP and DIP is further discussed in Section 4.3.

4.2.2.3 The Archival Information Package

An **Archival Information Package (AIP)** which is modeled in Figure 4-11 is a specialization of the IP which is defined to provide a concise way of referring to a set of information that has, in principle, all the qualities needed for permanent, or indefinite-long term, preservation of a designated Information Object. The AIP is itself an information object that is a container of other information objects. Within the AIP is the designated information object and it is called the Content Information.

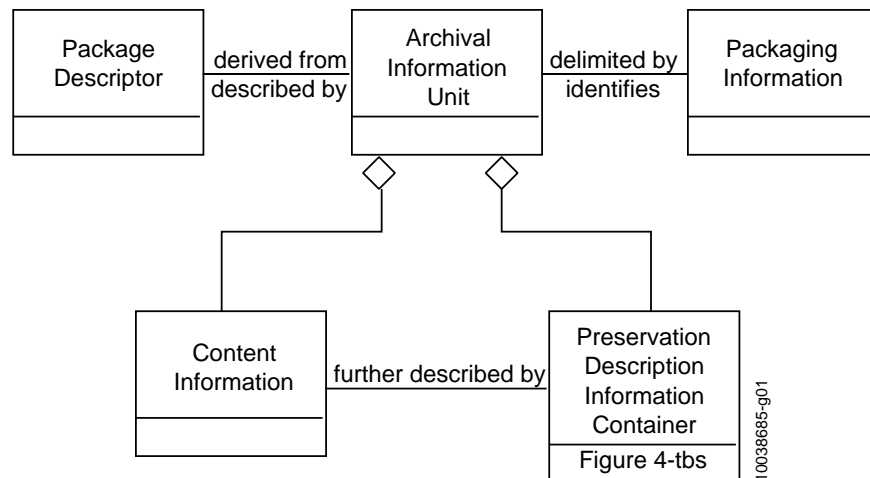


Figure 4-11. Archival Information Package(AIP)

Also within the AIP is an information object called the **Preservation Description Information Container (PDIC)**. The PDIC contains additional information about the Content Information and is needed to make the Content Information meaningful for the indefinite long-term.

The Preservation Description Information requirements in an AIP are must more stringent than the requirements for Preservation Description Information in an IP. While no PDI objects are mandatory in an IP, all PDI information must be present in an AIP. This is illustrated in Figure 4-12. In this case PDIC is used as a container object which contains all the PDI object types.

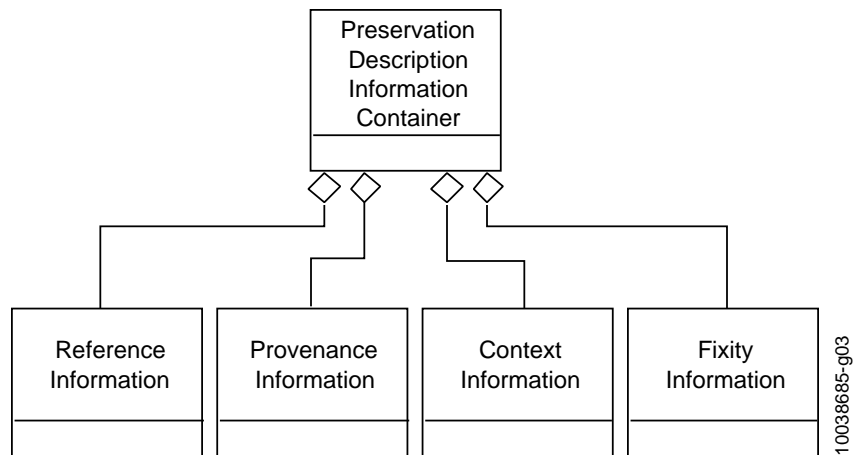


Figure 4-12. PDIC As a container for Preservation Description Information

The AIP is delimited and identified by the **Packaging Information**. The **Packaging Information** may actually be present as a structure on the media that contains the AIP or may be virtual in that it is contained in the OAIS Archival Storage function. However, the delimitation and internal identification functions must be well defined in an OAIS.

Each AIP is associated with a **Package Descriptor** which enables the consumer to locate information of potential interest, analyze that information, and order desired information.

Figure 4-13 gives a detailed view of the Archival Information Package by expanding the PDI and the Content Information. All the "contains" relationships discussed in this section are logical containment relationships. This type of containment relationship may be physical or may be accomplished via a pointer to another object in storage.

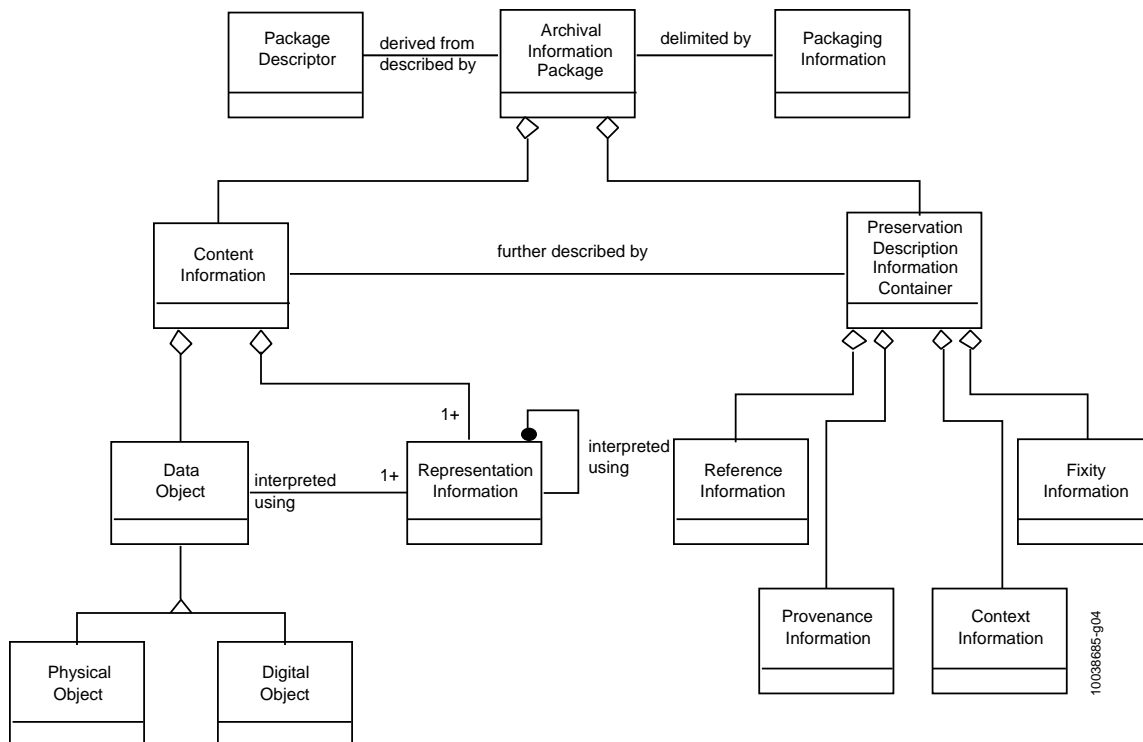


Figure 4 -13. Archival Information Package (Detailed View)

4.2.3.1. Specialization of the AIP and Associated Descriptors

Two important specializations of the AIP are discussed in this section, the **Archival Information Unit (AIU)** and the **Archive Information Collection (AIC)**. Figure 4-14 is an OMT diagram illustrating this specialization. Both AIU and AIC are subtypes of the AIP and as such contain constructs to enable both long term preservation and consumer access. The AIU represents the type used for the preservation function of a single content atomic object. The AIC organizes a set of AIPs (AIUs and other AIPs) along a thematic hierarchy which can support flexible and efficient access by the consumer community. Conceptually all the AIPs organized by an AIC are contained in the content field of that AIC. The differences between AIUs and AICs is the complexity of their Content Information and their associated Package Descriptors and Packaging Information.

This reference model considers the differences in the Content Information and associated Packaging and Descriptor functionality between AIU and AIC to be adequately complex and linked to justify the definition of separate object classes.

From an Access viewpoint, new subsetting and manipulation capabilities are beginning to blur the distinction between AICs and AIUs. Content objects which used to be viewed as atomic, can now viewed as containing a large variation of contents based on the subsetting parameters. chosen. In a more extreme example, the CI of an AIU may not exist as a physical entity. The CI could consist of several input files (or pointers to the AIUs containing these data files) and an algorithm which uses these files to create the data object of interest.

From an information preservation viewpoint the distinction between AIU and AIC remains clear. An AIU is viewed as having a single content information object which is described by exactly one PDIC. An AIC content object is viewed as a collection of other AICs and AIUs, each of which has its own PDIC. In addition, the AIC has its own PDIC which describes the collection criteria and process.

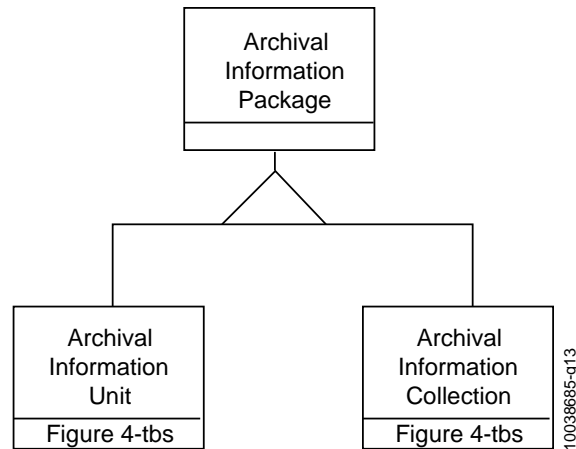


Fig 4-14. Archival specialization of the AIP

There are two specializations of the Package Descriptor, the Unit Descriptor and the Collection Descriptor. Figure 4-15 is an OMT diagram illustrates this specialization. The difference in these two classes is based on the functionality needed to effectively access the contents of an atomic AIU versus the functionality needed to effectively access to the contents of an AIC.

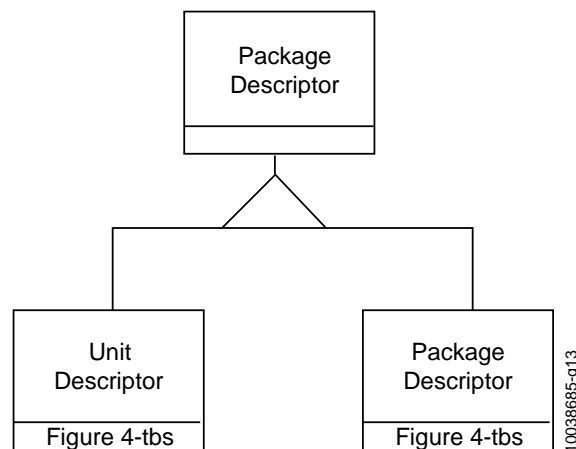


Fig 4-15. Archival specialization of the Package Descriptor

To aid in the understanding of these constructs the next two sections of this document will use an example of a company setting up an OAIS of digital versions of films. This example will focus on the information content of constructs in an AIP. Section 4.3 and the Illustrative Scenario in Section 7 will illustrate more of the details of the information transformations and dataflows in an OAIS.

4.2.3.1.1 Specializations of IOs within an AIP for atomic content objects

4.2.3.1.1.1 Archive Information Unit

The AIUs can be viewed as the "atoms" of information that the archive is tasked to store. A single AIU contains exactly one content information object (which may consist of multiple physical files) and exactly one set of PDI. When an information object is ingested into the OAIS a **Unit Descriptor**, which is a subtype of a Package Descriptor is created by extracting information from the CI and the PDI and appending it to the unique ordering information (e.g., cost and item number).

In the example of a digital film OAIS, the AIU for a single film can be viewed as three objects, one containing a digital image of the film in a proprietary format, one containing the Representation Information needed to understand the proprietary format (these two objects form the CI), and the other containing facts about the film such as date of creation, featured actors, director, producer, sequels, movie studio, and a checksum to ensure the integrity of the digital image (PDI). Since the OAIS reference model is implementation independent, these objects could be implemented as one file or multiple files. This type implementation dependent information is contained in the Packaging Information. When a movie is ingested into the OAIS a Unit Descriptor is created by extracting information from the CI and the PDI and appending it to the unique ordering information..

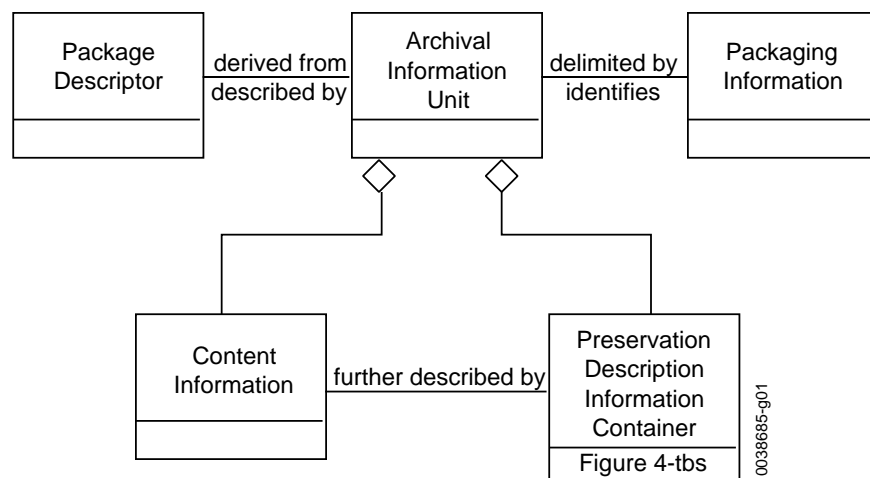


Figure 4-16. Archival Information Unit (AIU)

4.2.3.1.1.2 Unit Descriptor

The Unit Descriptor is a specialization of the Package Descriptor that always contains:

- a set of **Associated Descriptions** each of which describe the AIU content information from the point of view of a single **Access Aid**, and,
- a method for identifying and retrieving the AIU from Archival Storage

Figure 4-17 is an OMT diagram that illustrates the Unit Descriptor contents.

An important type of Access Aid is **Finding Aids** which are applications that assist the consumer in locating information of interest. A single AIU may have a number of Associated Descriptions which describe the CI using different technologies. Additionally as new description extraction and display technologies become available an archive may want

to update the Unit Descriptor associated with each of its AIUs to add a new Associated Description that utilizes the new technology to better describe the AIUs.

In the digital movie OASIS example, initially, there may be one Associated Description that is a free text description of a movie, another that is a five minute clip and another that is a row in a relational database that is used by film collectors to locate films of interest. After the archive has been operational for a period of time a technique for supplying compressed digital movies may be developed based on recording every tenth frame. The archivist may decide to create an additional type of Associated Description which is populated using the results of this new technique. If necessary, the user can run each of the AIUs contained in the archive through this compression technique and create a new Associated Description for each movie in the archive or simply include this Associated Description for new movies ingested into the OASIS.

Another important class of Associated Descriptions supply data for **Ordering Aids** that allow the consumer to discover the cost of and order AIUs of interest. The Ordering Aids also allow users to specify transformations to be applied to the AIUs prior to dissemination. These transformations can include data object transformations such as subsetting, subsampling or format transformations. The transformations can also involve modifying the PDI in the AIU prior to dissemination.

For example, the digital movie OASIS could allow a user to order a digital movie as a VHS tape, a laser disc or an MPEG object delivered on-line. Each of these would involve a format transformation and in theory an update to the PDI information in the AIP to create accurate PDI for the DIP.

In addition to the Associated Descriptions, the Unit Descriptor also contains **Access Methods** which enable authorized users to retrieve the AIU described by the Unit Descriptor. In most current archives, only internal archive processes and operations personnel are authorized to use these Access Methods.

For example, a pattern recognition technique might be created for digital movies and the digital movie OASIS might offer a service to search its archives for large structures such as the pyramids or a New York skyline. Note: the sort of service is very processing intensive, involving potentially large numbers of AIUs to be transferred from Storage to Dissemination and then running the appropriate process to analyze the CI from each AIU. If the results are generally useful, the archivist could summarize the results of this “content based query” into a new Associated Description. This technique is frequently referred to as data mining.

Although, the AIU and its associated Unit Descriptor provide the information necessary for a consumer to locate and order AIUs of interest, it can be impossible for a consumer to sort through the millions of Unit Descriptors in a large archive. This problem is addressed in the OASIS through the Archive Information Collection discussed in the next section.

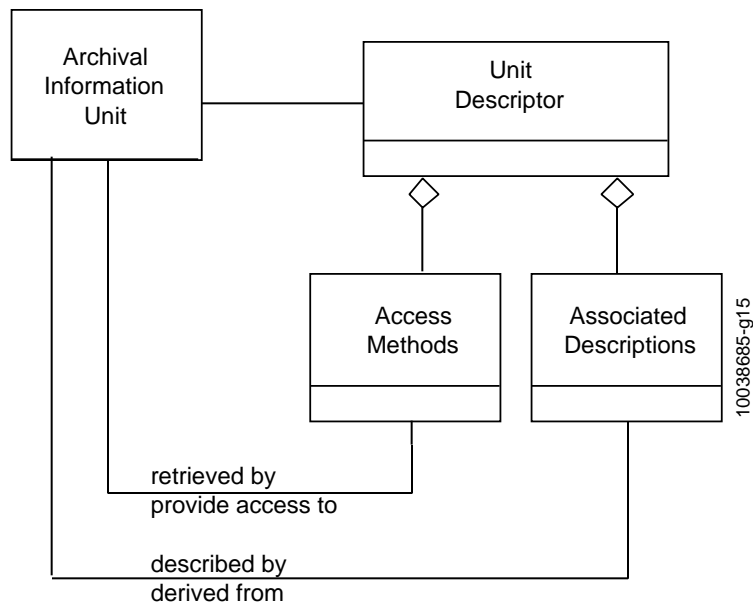


Figure 4-17 Unit Descriptor

4.2.3.1.2 Specialization of IOs in an AIP for compound content objects

4.2.3.1.2.1 Archive Information Collections

The content information of an AIC is composed of complete AIPs each of which have their own CI, PDI, and associated PI and PD. These AIPs are then aggregated into Archive Information Collections (AIC) using criteria determined by the archivist. Generally AICs are based on the AIUs of interest having common themes or origins and a common set of Associate Descriptions. At a minimum all OAIS can be viewed as having at least one AIC which contains all the AIPs held by the OAIS.

A logical model of a AIC is shown in Figure 4-18. As in the previous sections all the containment relationships are logical containment and may be physical or may be accomplished via a pointer to another object in storage. For example, the Content Information of an AIC can be created either by creating physical collections of the contained AIPs or by pointing to the contained AIPs. A single AIP can belong to any number of AICs.

For example the digital movies OAIS may have AICs based on the subject area of the movie such as mystery, science fiction, or horror. In addition the archive may have AICs based on other factors such as director or lead actor.

An important feature of the AIC as shown in is the fact that an AIC is a complete AIP which contains PDI which provides further information about the AIC such as provenance on when and why it was created, context to related AICs, and fixity information. This is in addition to the PDI contained in member AIPs. This type of information is often necessary for a consumer to have confidence in the reliability of an AIC. In the digital movies OAIS example, the usefulness of an AIC of movies starring John Wayne is to some extent based on the provenance of when the collection was created or last updated.

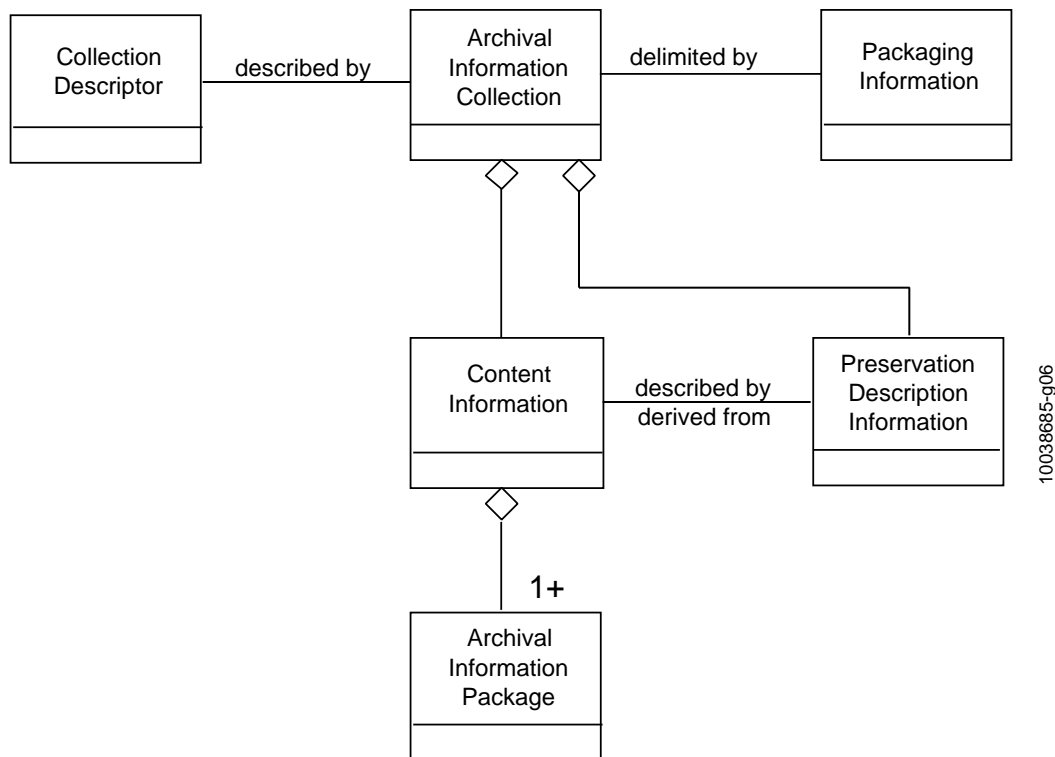


Figure 4-18. Archive Information Collections Logical View

4.2.3.1.2.2 Collection Descriptors

The **Collection Descriptor** is a subtype of the Package Descriptor which has added structures to better handle the complex content information of an AIC. The Collection Descriptor, which is modeled in Figure 4-19 contains the information classes which are contained in the Unit Descriptor. The Access Methods of a Collection Descriptor provide a user with access to the entire Content Information of the associated AIC and the PDI for the AIC not for members of the AIC.

There are two types of Associated Descriptions in an Collection Descriptor:

- Associated Description that describe the collection as a whole (called Collection Description in Figure 4-19)
- Associated Description that separately describe each member of the collection (called Member Description in Figure 4-19)

the Finding Aid would allow the consumer visibility of these future AIPs but a the Associated description for the Ordering Aid and/or Access Methods would contain the information the this product was not currently available.

4.2.3 Data Management Data

Currently, Package Descriptors are stored in persistent storage such as database management systems to enable easy, flexible access and update to the contained Associated Descriptions. In addition to the Package Descriptors discussed in the previous sections, all the information needed for the operation of an archive would be stored in databases as persistent data classes. Some examples of this data are accounting data for customer billing and authorization, policy data, subscription data for repeating requests, and statistical data for generating reports to archive management. These classes are intended as examples rather than an exhaustive list of the data required for archive administration. Figure 4-18 is an OMT diagram which illustrates the various types of "data management data" within the OAIS

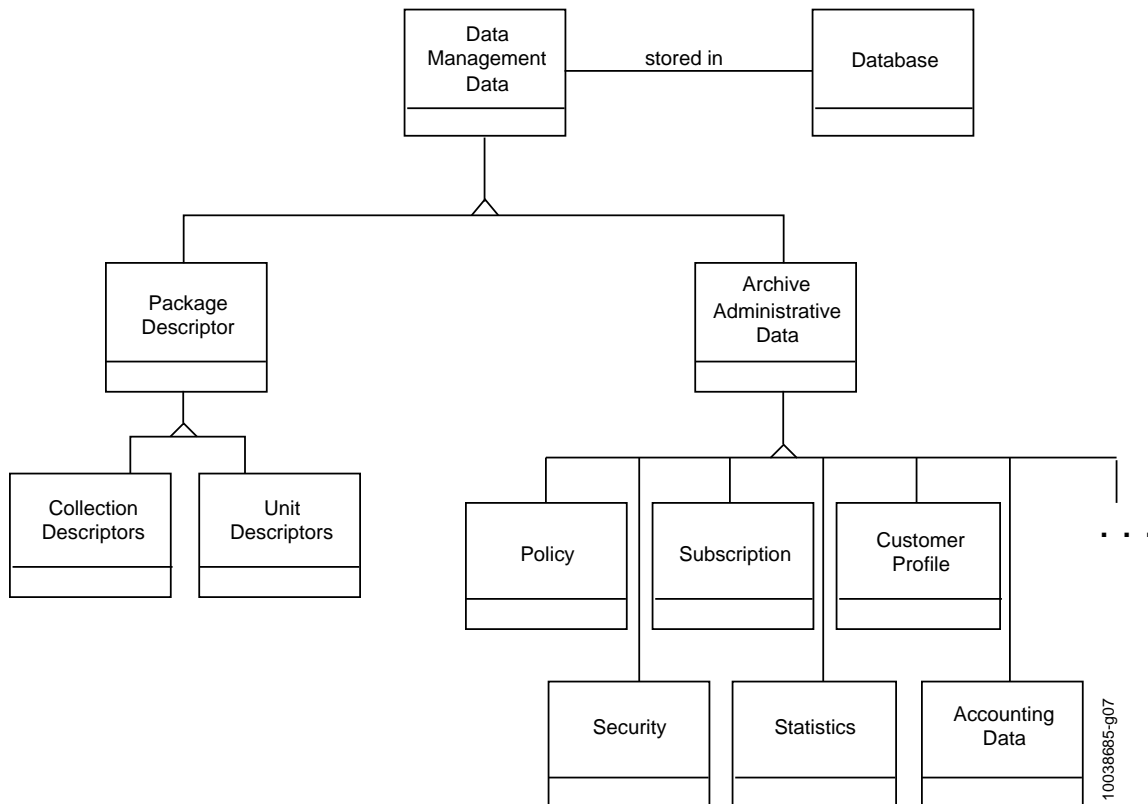


Figure 4-18. Data Management Data

Stuff for Annex D

The modification to OMT includes the ability to annotate two way relationships between two Object classes. This is accomplished by reading text above a relationship line from left to right and the text below a relationship from right to left.